基于信息熵和用户行为一致性的协同过滤分组推荐 *

苏梦珂，杨煜普

(上海交通大学 自动化系 系统控制与信息处理教育部重点实验室，上海 200240)

摘　要：在仅以输入评分矩阵作为唯一算法输入的协同过滤推荐算法研究中，针对数据的质量不同带来的差异性对推荐结果的影响这一问题，包括对数据质量方面的重视与关注、如何刻画质量差异性以及如何针对不同质量数据的用户组别进行分组推荐建模等问题，提出针对数据质量的刻画，综合考虑用户行为一致性和用户信息熵两个指标对数据质量进行评价并对用户进行分组。对于不同组别的用户在分析其历史行为的基础上可以进行更精准的推荐建模。实验结果表明，数据质量的差异性确实对推荐精度的提升有着重要的影响，同时论证了对用户进行分组推荐的必要性。实验结果同时表明，运用用户行为一致性和用户信息熵两个指标的综合刻画带来的精度提升效果最为显著。

关键词：信息熵；噪声刻画；数据质量差异性；用户行为一致性；协同过滤

中图分类号：TP311　　doi: 10.3969/j.issn.1001-3695.2018.05.0391

Collaborative filtering group recommendation based on information entropy and user behavior consistency

Su Mengke, Yang Yupu

(*Key Laboratory of System Control & Information Processing*, *Ministry of Education of China*, *Dept. of Automation*, *Shanghai Jiao Tong University*, *Shanghai* 200240, *China*)

Abstract: For the scoring matrix as the unique algorithm input of the collaborative filtering recommendation algorithm, the differences in the quality of the data have great impact on the recommendation results, including arousing the attention to data quality, how to characterize quality differences, and how to group users and recommend on the basis of user groups with different quality data. This paper proposes a description of data quality, comprehensively considers the "user behavior consistency" and "user information entropy" to evaluate the data quality. Users of different groups can perform more accurate recommendation results based on analyzing their historical behavior. The experimental results show that the difference in data quality does have an important impact on the improvement of recommendation accuracy, and at the same time demonstrate the necessity of group recommendation. The experimental results also show that the accuracy of the combination of the two aspects of "user behavior consistency" and "user information entropy" is the most significant.

Key words: information entropy; noise description; data quality difference; user behavior consistency; collaborative filtering

0　引言

互联网技术的快速发展加速了网络资源的急剧膨胀，每天都有大量的信息充斥在网络中，这种膨胀的信息过载问题使得用户通过传统的检索模式去寻找自己感兴趣的信息的代价越来越高，同时用户也很难对爆炸式增长的信息进行有效的处理和利用。推荐系统作为当下社会解决信息过载问题的一项重要技术，已经被广泛地应用在各大平台如亚马逊电子商务平台和一些社交网站平台。其中基于协同过滤[1]的推荐算法自出现以来就得到了广泛的研究和应用，主要由于其实现的简单性和易扩展性。经典的基于用户的协同过滤算法[2]主要是为每个用户找到一个相似度高的用户集合，利用两个用户之间存在共同评分项目计算用户之间的相似性。而基于模型[3,4]的协同过滤推荐算法是利用评分数据对用户评分规律进行数学建模,其中矩阵分解模型[5]把用户和项目映射到共同的低维隐含空间，并尝试通过用户和项目在隐空间的向量积解释评分，后来随着 Netflix 大赛的发展，矩阵分解及其改进模型因突出表现脱颖而出。

基于模型的协同过滤算法的优点在于可以提高推荐的准确性，算法的关键在于利用训练数据集离线学习一个预测模型，而算法和数据是影响这个模型准确性的两个相当重要的因素。推荐算法是基于数据集提供的数据质量不存在差异，基于不同用户的行为数据都准确地代表了用户的真实喜好，继而在建模

收稿日期：2018-05-16；**修回日期：**2018-07-06　　**基金项目：**国家自然科学基金资助项目（5177070084）

作者简介：苏梦珂（1994-），女，硕士，主要研究方向为推荐系统、数据挖掘、机器学习（sumengke2016sjtu@ sjtu.edu.cn）；杨煜普（1957-），男，教授，博导，主要研究方向为智能控制、机器学习.

时对所有用户进行同等看待不加分类。但是在现实应用中的推荐系统如电子商务平台会存在一些恶意用户给出参考价值很小的用户反馈,而有些用户的反馈会十分有利于对他们进行推荐。即不同用户的行为数据在质量方面存在差异性，有些用户的行为前后比较一致，反馈比较稳定，利用这些数据不仅降低建模难度，并可以给这些用户提供更精准的结果，而有些用户的行为前后反差较大，对模型来说这些数据较难利用且结果也不够准确。所以数据的质量和数量极大地影响了推荐结果的准确性[6]，即数据中的噪声问题，而数据中的噪声问题却一直以来没有得到广泛的关注。推荐系统的噪声一般可以分为两类：a)为了某种目的，为了提高商业利益或者恶意扰乱而出现的蓄意噪声[7]；b)用户打分太过随意，并没有真实的表达自己的想法的自然噪声[8]。

近来的相关研究只是单纯地把数据质量问题转换为部分去噪问题,如 Chirita 等人[9]提出了一种识别恶意用户的评价指标，Bilge 等人[10]为了识别恶意用户和水军账号，采用 K-均值聚类算法划分不同的用户组别，Cao 等人[11]提从半监督(semi-shilling attack detection，Semi-SAD)的角度出发，利用少量数据预训练一个贝叶斯分类器再自适应于所有数据得到最终的分类器。这些方法虽然使得噪声引起的推荐精度问题得到了解决，但是共同弊端在于需要训练复杂的模型，并且调参复杂，而噪声只是数据质量的一种体现。

刘江冬等人[16]提出借鉴信息熵的概念，综合考虑用户信息熵和评分时效性过滤部分用户，从而提高推荐的准确性。于鹏华[17]从数据的角度综合考量数据的质量和数量问题，对评分数据进行分组,分析了数据的质量和数量差异对推荐结果的影响。张佳等人[18]借鉴用户信息熵来表达用户的评分分布，并确定评分倾向性程度，在传统的基于用户的协同过滤算法中确定某一用户的最近邻时利用信息熵将某些明显倾向不同的用户剔除，提高了最后的推荐精度。高翠华等人[19]综合考虑信息熵和模糊聚类，利用信息熵衡量隶属度的不确定性，提出融合信息熵加权的模糊聚类协同过滤算法，在提高推荐精度的同时简化了算法的复杂度。Kluver 等人[8]提出可以用信息熵刻画用户的评分质量，Bellogín 等人[12]通过分析用户历史行为提出一种新的评分质量评价。为了充分说明数据质量对推荐结果的影响，本文综合考虑信息熵和用户行为一致性来共同全面地刻画不同质量的数据，并提出基于不同质量和数量的数据子集的分组协同过滤推荐算法。本文提出综合考虑用户评分两种极端情况随意和集中来分析用户质量，将用户分为不同质量的数据子集并进行分组推荐，进一步实现了针对不同用户的个性化建模，并提高了推荐精度。

1 问题基本描述

1.1 基本模型 BMF

矩阵分解模型的输入如表 1 所示。表 1 中显示的是 m 个用户对 n 个项目的所有的评分集合，其中如果用户 u 对某个项目 i 进行了观影评价，则 r_{ui} 表示相应的评分，一般的评分矩阵比较稀疏，用户未评论过的电影在矩阵内评分用 0 表示。如何利用已有的评分数据通过数学建模的方式将用户和项目潜在的评分规则公式化,并对未知评分进行预测是进行推荐的关键步骤。

表 1 评分矩阵

用户	项目			
	i_1	i_2	…	i_n
u_1	r_{11}	r_{12}	…	r_{1n}
u_2	r_{21}	r_{22}	…	r_{2n}
…	…	…	…	…
u_m	r_{m1}	r_{m2}	…	r_{mn}

矩阵分解 BMF 模型是隐语义模型 (latent factor model) [13]的改进模型，因为其在推荐系统数据比赛 Netflix 比赛 2009 年的杰出表现而成为最热门的推荐算法。矩阵分解模型 BiasedMF(Biased matrix factorization)算法通过寻找一个低维的隐含因子空间，把原始用户和项目映射到这一低维空间，项目 i 映射成向量 q_i，向量分量表示项目 i 对这些基本因子的包含程度，用户 u 映射成向量 p_u，向量分量是用户对该因子的偏好程度的表征。经过低维映射后，某用户 u 对某项目 i 的偏好程度可以用用户和物品向量的内积 $q_i^T p_u$ 来表示。

基于矩阵分解的推荐算法以所示矩阵作为算法的输入，通过挖掘用户和项目的隐含潜在因子进行数学建模，并对用户未知的评分进行预测。本文选用 BiasedMF(BMF)算法为基础，融合用户的信息熵和行为一致性提出了改进方案。BMF 的基本算法如下：

$$b_{ui} = \mu + b_u + b_i \tag{1}$$

$$\hat{r}_{ui} = b_{ui} + q_i^T p_u \tag{2}$$

$$\min_{p_u, q_i} \sum_{(u,i)\in R_{new}} \left[(\hat{r}_{ui} - r_{ui})^2 + \lambda(\|q_i\|^2 + \|p_u\|^2 + {b_u}^2 + {b_i}^2) \right] \tag{3}$$

式(1)表示评分的偏置项构成，其中 μ 表示整体平均值，b_u 表示用户偏置，b_i 表示项目偏置；式(2)表示预测评分由模型预测结果和偏置项构成。训练模型使得 $\hat{r}_{ui}$ 无限接近于 r_{ui}，即转换成式(3)所示的最优化问题，其中 λ 表示正则化项，目的是为了提高模型的泛化能力。选择常用的随机梯度下降法寻找到最优的 P 和 Q。模型训练好以后，就可以对用户未评分电影进行预测，并按照评分从高到低的集合进行推荐。

1.2 噪声问题

大多相关学者的研究工作都是收集到原始的评分矩阵，根据某种策略将数据分为训练集和测试集，然后进行建模与测试。因为评分矩阵作为协同过滤算法的唯一输入，所以评分矩阵的质量差异性会在很大程度上影响算法的最终结果。如何刻画不同用户的评分质量，并对用户进行分组推荐是本文关心的问题。

为了分析评分质量的差异与推荐精度的关系，针对不同的用户存在不同程度的噪声数据的问题，本文针对数据中的噪声问题，综合考虑用户信息熵和用户历史行为来刻画用户的评分

质量差异，并将用户分为不同的质量子集。根据文献[8]指出的可信度低的用户的评分存在过于集中，如评分一边倒或者只是针对某些特定的目标的问题，本文引入信息熵来刻画用户评分质量。信息熵可以表征随机变量的取值不确定性的情况，刻画一个随机变量的概率分布情况，随机变量的信息熵的值与其分布的混乱程度成正比。设某一变量的信息熵定义如式(4)所示。

$$H(Y) = -\int f(y) lbf(y) dy \tag{4}$$

由式(4)可知，信息熵与变量的概率分布密切相关，用户信息熵可以反映用户评分的丰富程度，如果用户信息熵过低，即可表示用户的评分过于集中。

但是如果只是过滤掉集中性过强的用户，那些打分丰富的用户的并不一定都是可靠的评分，所以不能单独以集中性来刻画用户的评分质量。根据 Bellogin 提出的基于用户历史行为数据引入用户行为一致性[12]来从另一个角度衡量用户的评分质量。用户行为一致性反映了用户的评分是否前后连贯，可根据这一指标将用户划分为行为一致性程度较高和较低的不同组别。综合考虑用户信息熵和用户行为一致性分析不同的数据质量对推荐结果的影响。

2　模型改进

2.1　评分的质量刻画

2.1.1 引入用户信息熵刻画用户评分质量

如何对评分矩阵的质量差异性进行刻画，如何将不同的用户分为不同的质量子集，本文从数据的噪声的角度来刻画数据的质量，并提出综合分析用户的信息熵和历史行为来分析用户评分的质量。根据式(4)的定义，假设 $\{R_u = r_1, r_2, \ldots r_i \ldots\}$ 表示用户 u 的所有评分信息。对用户 u，定义用户评分取值的概率为该评级出现的次数与评分总次数的占比，即 P_{ug} 如式(5)所示，用户的信息熵定义为 $C_1(u)$ 如式(6)所示。

$$P_{ug} = \frac{\left[\sum_{r_i \in R_u} I(r_i = g)\right]}{|R_u|}; g = 1,2,3,4,5 \tag{5}$$

$$C_1(u) = \sum_{g=1}^{5} -P_{ug} lb(P_{ug}) \tag{6}$$

用户信息熵反映了用户评分偏水军特点的可能性。根据文中公式定义，用户信息熵偏低则表明用户偏水军的可能性越高。

2.1.2 基于用户历史行为刻画用户评分质量

基于用户的历史评分行为，如图 1 表示两个用户对类似题材分布下的电影的评分集合，基于用户的评分高低反应用户的好恶，图 1 左的用户相对于爱情和伦理片更喜欢惊悚类的电影，而图 1 右的用户的行为数据比较分散，不易挖掘规律。长远看来有理由相信图 1 中的用户（左）评分相对于用户（右）比较稳定，因为用户（左）对于相似类型的电影评分也比较相似。

假设用户行为前后连贯表现在对同种类型或者相近类型地评分偏差较小，利用用户在相似物品空间的评分偏差来定义用户行为的稳定性。对任一用户有过评分行为的项目划分为不同的特征空间，假设 $R(u, f)$ 表示用户对特定特征空间的物品的评分集合，$\bar{r}_{uf}$ 表示用户在这一特征下的评分平均值，$R(u, F)$ 表示用户 u 的所有评分集合，则用户行为一致性 $C_2(u)$ 可以用用户在各种特征空间的评分下的偏差来表征，如式(8)所示。

$$C_{2f}(u) = \sqrt{\sum_{i \in R(u,f)} (r_{ui} - \bar{r}_{uf})^2} \tag{7}$$

$$C_2(u) = -\sum_{f \in F} C_{2f}(u) \frac{\|R(u,f)\|}{\|R(u,F)\|} \tag{8}$$

其中：$C_{2f}(u)$ 表示用户的某种特定评分方差；$C_2(u)$ 表示用户整体评分方差，实质为用户的特定评分方差的加权平均，其值与用户行为一致性成正比关系，用户行为前后行为一致，则这部分用户较容易建模。

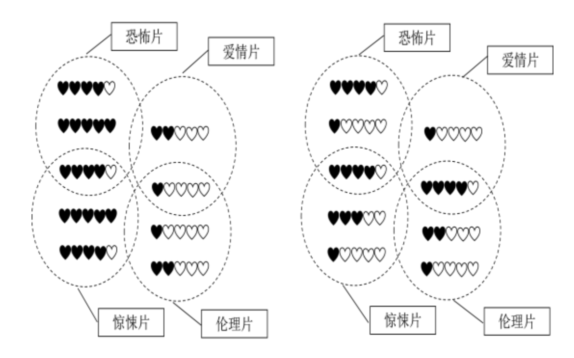


图 1　用户行为一致性

2.2　用户分组并进行分组推荐

本文对所提的 $C_1(u)$ 和 $C_2(u)$ 采用顺序式用户分组，通过对用户的评分数据分析，每个用户得到相应的 $C_1(u)$ 和 $C_2(u)$。对于信息熵这个评价指标，由于信息熵是针对用户评分数据具有一边倒的特点而定义的，所以首先根据 $C_1(u)$ 将用户信息熵较低的用户采取直接过滤的方式，即在原始数据中保留评分质量良好的大部分数据；然后根据 $C_2(u)$ 的值将用户聚类为困难用户和容易用户，分析不同组别的质量差异性对推荐精度的影响。

3　实验与分析

3.1　实验数据集

为了测试数据的质量差异性对最终推荐精度的影响，以及验证质量指标的有效性和分组推荐的必要性，利用著名的 Movielens1M（ml-1m）电影评分公开真实数据集[14]对本文提出的算法进行实验评估。该数据集包括 6 000 多用户的评分数据 1 000 209 条，每个用户评论的电影在 20 部以上，一共有 3 900 部电影。其中评分值反映了用户对电影的喜爱程度，评分值越大代表用户对电影评价越高。该数据集属于数据量易处理里涵盖信息比较丰富的数据集，在用户的数据质量方面存在差异性，适合进行此次验证。

3.2　评价指标

不同的评价指标适用于不同的研究环境。由于本文主要是

针对数据质量差异性进行刻画，为了体现不同质量数据组对推荐精度的影响，并分析针对不同质量差异性造成的推荐精度的影响，本文采用评分预测中的均方根误差(RMSE)[15]指标作为此次实验的评估。RMSE 是通过利用在训练集上得到的模型对未知评分进行预测后计算预测的用户评分和实际的用户评分之差来评估推荐精度。RMSE 因其直观表征推荐的精度而被广泛采用。RMSE 与推荐精度成反比，如式(9)和(10)所示。

$$RMSE(u)=\sqrt{\frac{\sum_{i=1}^{n}(r_{ui}-\hat{r}_{ui})^2}{n}} \tag{9}$$

$$RMSE=\frac{\sum_{u=1}^{m}RMSE(u)}{m} \tag{10}$$

3.3 实验步骤

3.3.1 刻画用户评分质量

对数据集中的用户作为基准，对于每一条用户的评分数据，计算相应的用户信息熵和用户行为一致性，得到信息熵 $C_1(u)$ 如图 2、3 所示，一致性 $C_2(u)$ 如图 4、5 所示。

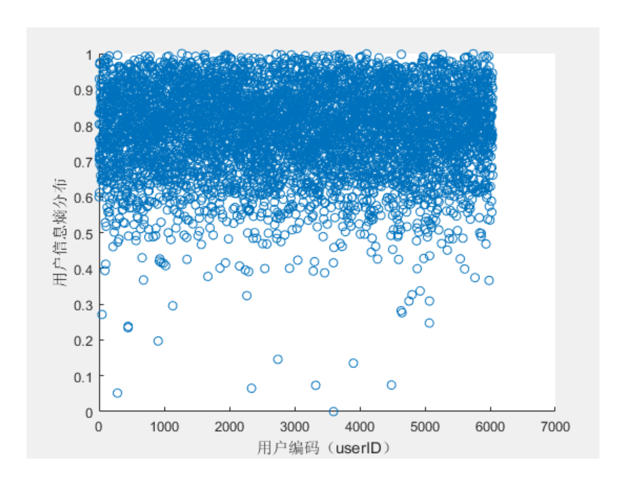


图 2　用户信息熵(归一化之后)

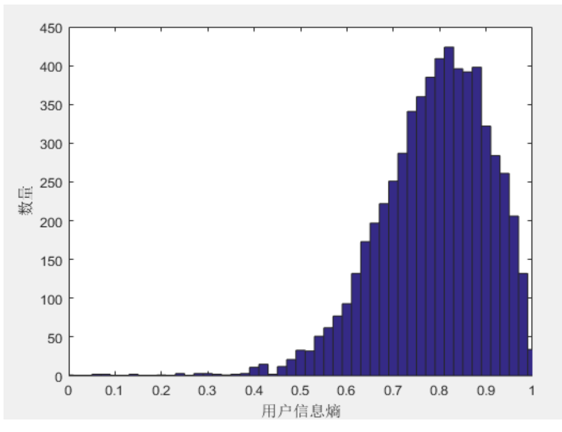


图 3　用户信息熵数量统计

图 2 和 3 分别表示对 6 040 个用户的信息熵的分布和数量统计，大部分的评分质量是可靠的。$C_1(u)$ 取不同阈值，通过取阈值为 $C_1(u)=\{0.1,0.2,\ldots 0.8\}$，发现取 $C_1(u)=0.5$ 时精度提升最大，之后随着 $C_1(u)$ 的增大，会使评分矩阵越来越稀疏，从而也影响推荐的精度。

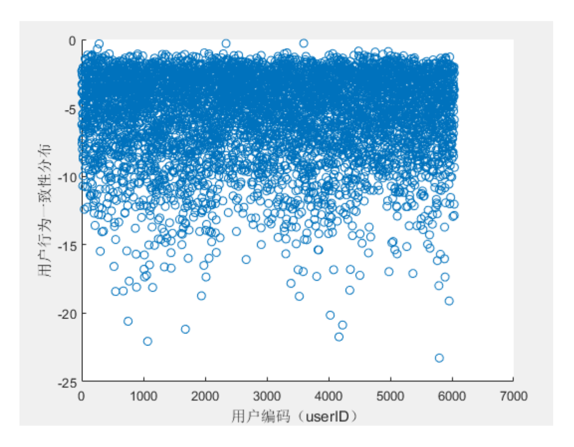


图 4　用户行为一致性

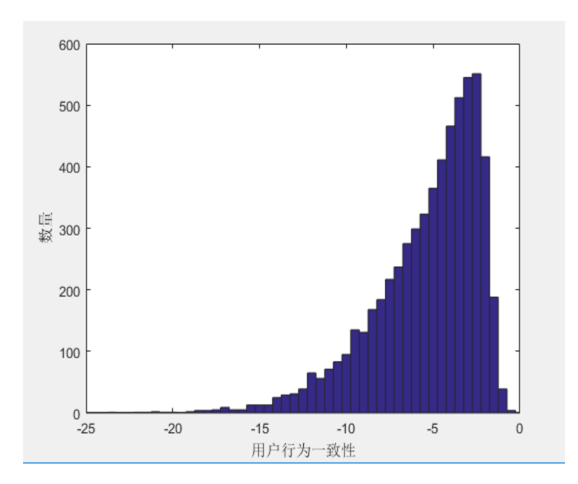


图 5　用户行为一致性数量统计

3.3.2 用户分组推荐

首先对用户根据 $C_1(u)$ 的值，选取不同的阈值进行数据过滤。对于低于阈值的用户评分数据，采取直接删除这一部分用户的做法。实验表明 $C_1(u)=0.5$ 的情况下，精度提升最大。

在此基础上根据 $C_2(u)$ 的值将用户分为困难用户和容易用户（取 $C_2(u)=-8$），此时，困难用户组评分数量为 509 623，容易用户组的评分数量为 490 586，两组分别接近原始数据集的 50%，以排除数据数量对推荐结果的影响。两组用户都采用 5 折交叉验证划分为数据集和训练集，分别表示为 tr_e、te_e、tr_d、te_d。其中 $s_1=(tr_d,te_d)$ 表示对困难用户组建模推荐，$s_4=(tr_e,te_e)$ 表示对容易用户组建模推荐，$s_2=(tr_e\cup tr_d,te_e\cup te_d)$ 表示原始数据建模推荐（对比的基准），s_3 表示对原始数据进行信息熵过滤之后的用户组进行建模推荐。图 6 表示不同质量分组的情况下的误差 RMSE，其中横轴 $(1,2,3,4)$ 分别表示 (s_1,s_2,s_3,s_4)。

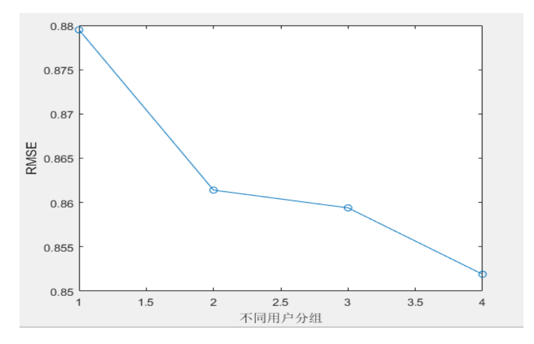


图 6　不同质量情况下的误差 RMSE

3.3.3 对比实验

本文提出的融合用户信息熵和用户行为一致性改进的基于模型的协同过滤算法(ABMF)对比基本的 BMF 算法，RMSE 指标降低了 1.1%。文献[16]提出的 UEITMF 方法融合信息熵和时效性的改进，考虑了实时动态这一特性，但是相较原始方法精度提升不明显，而且增加了算法的复杂度。文献[17]只是从数据质量的角度考虑，单一地考虑了用户前后行为的变化，推荐精度也有一定提升。Entropy-based-CF[18]是在基于用户的协同过滤的基础上融合信息熵，算法的逻辑性和理解性较高。不同算法的 RMSE 对比如表 2 所示。

表 2　不同算法的 RMSE 对比

算法	BMF	ABMF	UEITMF[16]	文献[17]	Entropy-based-CF[18]
RMSE	0.861	0.852	0.859	0.855	0.865

3.4　实验结果分析

对比 s_2 和 s_3，不考虑用户行为一致性，只对用户进行信息熵过滤，虽然取 $C_1(u)=0.5$ 的情况下，推荐的精度由 0.861 4 降低到 0.859 4，效果微弱，但是这一质量指标的提出对于大数据情况下分析如电子商务系统中刷单的行为应该会具有明显的效果。

经过过滤质量极差的用户后，针对剩余的用户根据进行 $C_2(u)$ 的值分组，如图 6 所示，s_1 表示在困难用户子集上建模，并在困难用户子集上测试；s_4 表示在容易用户子集上建模，并在容易用户子集上测试；s_3 相对于 s_2 推荐精度提升了 0.2%；s_4 相对于 s_2 推荐精度提升了 1.1%，s_4 相对 s_1 推荐指标变化了 3.2%。推荐精度的变化说明进行分组推荐是十分必要的，并且 (s_1, s_2, s_3, s_4) 分别表示数据的质量从低到高，明显 RMSE 呈下降趋势，即推荐精度呈提高趋势。实验结果论证了本文提出的信息熵和用户行为一致性综合分析用户评分质量的指标是十分有效的，不同质量组的建模推荐精度存在着显著差异。

4　结束语

本文从一个新型的角度即数据质量出发，提出提高基于模型的推荐算法精度的研究，综合评分集中性和随意性，分别通过用户信息熵和行为稳定度来刻画用户评分的质量差异性，并提出基于不同数据质量把用户分为不同组别并进行分开建模的改进方法。实验结果充分表明了数据质量的差异性对推荐精度的提高有重要的影响，所以在当下大数据的发展环境下，数据的质量问题应该引起广泛关注。数据质量问题会随着数据规模的增大凸显，数据集越大数据质量差异性会更显著，如何利用信息熵和用户行为一致性这两个指标对大数据集进行更好的分析，并对用户进行合理分组是接下来的研究方向。本文针对用户实现了评分质量分组，如何针对不同项目设置合理的分类标准，从不同角度定义相应的质量指标对数据进行分类，也是未来考虑的方向。

参考文献：

[1] Chu Wei, Park S T. Personalized recommendation on dynamic contents using predictive bilinear models [C]// Proc of the 18th International Conference on World Wide Web. New York: ACM Press, 2009: 691-700.

[2] 沈键，杨煜普．基于二阶段相似度学习的协同过滤推荐算法 [J]．计算机应用研究, 2013, 30 (3): 715-719. (Shen Jian, Yang Yupu. Collaborative filtering recommendation algorithm based on two stages of similarity learning [J]. Application Research of Computers, 2013, 30 (3): 715-719.)

[3] 吴金龙．Netflix Prize 中的协同过滤算法 [D]．北京：北京大学，2010. (Wu Jinlong. Collaborative filtering algorithm in the Netflix Prize [D]. Beijing: Beijing University, 2010.)

[4] Robert B, Yehuda K, CHRIS V. Modeling relationships at multiple scales to improve accuracy of large recommender systems [C]// Proc of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2007: 95-104.

[5] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems [J]. Computer, 2009, 42 (8): 30-37.

[6] 孟祥武，刘树栋，张玉洁，等．社会化推荐系统研究 [J]．软件学报，2015, 26 (6): 1356-1372. (Meng Xiangwu, Liu Shudong, Zhang Yujie, *et al*. Research on social recommender systems [J]. Journal of Software, 2015, 26 (6): 1356-1372.)

[7] Gunes I, Kaleli C, Bilge A, et a1. Shilling attacks against recommender systems: a comprehensive survey [J]. Artificial Intelligence Review, 2014, 42 (4): 767-799.

[8] Kluver D, Nguyen T, Ekstrand M, et a1. How many bits perrating [C]// Proc of the 6th ACM Conference on Recommender systems. Dublin, Ireland: ACM Press, 2012: 99-106.

[9] Chirita P A, Nejdl W, Zamfir C. Preventing shilling attacks in online recommender systems [C]// Proc of the 7th Annual ACM International Workshop on Web Information and Data Management. New York: ACM Press, 2005: 67-74.

[10] Bilge A, ZdemiRZ, Polat H. A novel shilling attack detectionmethod [J]. Procedia Computer Science, 2014, 31: 165-174.

[11] Cao Jie, Wu Zhiang, Mao Bo, *et al*. Shilling attack detection utilizing semi-supervised learning method for collaborative recommender system [J]. World Wide Web, 2013, 16 (5//6): 729-748.

[12] Bellogin A, Said A, Pdevries A. The magic barrier of recommender systems no magic just ratings [C]// Proc of User Modeling, Adaption, and Personalization. Aalborg, Denmark: SpringerInternational Publishing, 2014: 25-36.

[13] Hofmann T. Latent semantic models for collaborative filtering [J]. ACM Trans on Information Systems, 2004, 22 (1): 89-115

[14] 张学胜．面向数据稀疏的协同过滤推荐算法研究 [D]．合肥：中国科学技术大学，2011. (Zhang Xuesheng. Research on collaborative filtering recommendation algorithm for data sparse [D]. Hefei: University of

Science and Technology of China, 2011.)

[15] Pang Yanwei, Ma Zhao, Pan Jing, *et al.* Robust sparse tensor decomposition by probabilistic latent semantic analysis [C]// Proc of the 6th International Conference on Image and Graphics. Washington DC: IEEE Computer Society, 2011: 893-896. 刘江冬，梁刚，冯程，等. 基于信息熵和时效性的协同过滤推荐 [J]. 计算机应用 2016, 36 (9): 2531-2534. (Liu Jiangdong, Liang Gang, Feng Cheng, *et al.* Collaborative filtering recommendation based on information entropy and timeliness [J]. Journal of Computer Applications 2016, 36 (9): 2531-2534.)

[16] 于鹏华. 数据数量与质量的推荐系统若干问题研究 [D]. 杭州：浙江大学, 2016. (Yu Penghua. Research on several problems of recommendation system for data quantity and quality [D]. Hangzhou: Zhejiang University, 2016.)

[17] 张佳，林耀进，林梦雷，等. 基于信息熵的协同过滤算法 [J]. 山东大学学报：工学版, 2016, 46 (2): 43-50. (Zhang Jia, Lin Yaojin, Lin Menglei, *et al.* Collaborative filtering algorithm based on information entropy [J]. Journal of Shandong University: Engineering Edition, 2016, 46 (2): 43-50.)

[18] 高翠芳，黄珊维，沈莞蔷，等. 基于信息熵加权的协同聚类改进算法 [J]. 计算机应用研究 2015, 32 (4): 1016-1018. (Gao Cuifang, Huang Shanwei, Shen Wanqiang, *et al.* Improved algorithm for collaborative clustering based on information entropy weighting [J]. Application Research of Computers, 2015, 32 (4): 1016-1018.)